An Improved MIP Formulation for the Unit Commitment Problem

Semih Atakan\textsuperscript{1}, Guglielmo Lulli\textsuperscript{2}, Suvrajeet Sen\textsuperscript{1}

\textsuperscript{1}Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, USA
\textsuperscript{2}Dipartimento di Informatica Sistemistica e Comunicazione, Università di Milano-Bicocca, Milan, Italy

August 31, 2015

Abstract

In this paper, we present an MIP formulation for the Unit Commitment problem that leads to significant computational time savings and almost-integral solutions when compared to the state-of-the-art formulations in the literature. Using a variety of test instances from the literature, we provide empirical evidence that the polyhedral structure of the new formulation is tighter than the polyhedral structure of the state-of-the-art formulation. An important consequence of the new formulation is that it allows us to solve real-life instances of the problem in less than ten minutes on an ordinary desktop PC. In comparison, the state-of-the-art formulation is unable to provide a provably optimal solution in two hours of computing time on the same machine.

1. Introduction

Every day regional electricity networks deliver millions of kilowatt-hours (kWh) of energy from generating units to consumers. These production requirements vary by season, day-of-the-week, and hour. As a result, efficient scheduling of electricity production continues to attract significant attention from both industry and academy in the form of the so-called unit commitment (UC) problem. Such a model must recommend which generators to use, and how much they should produce so that demand over a planning horizon (say a week) is met, while obeying certain operating rules, maintenance schedules, and in some instances, even transmission capacity requirements. The goal is to obtain the most cost-effective operating schedule over a large set of generators, and ensure that the demand is completely fulfilled over the planning horizon.

With the liberalization of the energy industry, the introduction of energy market concepts and privatization, the role of the UC models has changed (see Hobbs et al., 2001). UC problems are solved by a variety of constituents of the electric power industry, each for a different purpose. A specific utility, e.g., the Los Angeles Department of Water and Power, solves the UC problem for the purposes of planning production within their delivery area. Their costs also helps them to place bids in the California Independent System Operator (CAISO) exchange market. On the other hand, the Independent System Operator (ISO) solves its UC model in order to decide which bids to accept and to set prices that will be paid to the suppliers. These problems are much larger because they involve multiple suppliers, each sending in their bids, and with a number of generating units in the order of thousands. Providing high-quality solutions to these real-life problems is computationally very demanding but has the potential for notable reduction in costs. Even a small improvement can lead to significant savings in a billion-dollar market. A 2011 report of the Federal Energy Regulatory Commission (FERC) suggests that savings approaching a $100 million annually can be expected by replacing heuristics with methods that seek optimal solutions (see O’Neill et al. (2011)). Consequently, developing solution methods that can achieve high-quality solutions in a limited amount of time has been the focus of significant research over the last several decades.

Many optimization methods have been proposed to solve the UC problem. For example, we mention branch-and-bound methods Turgeon (1977), dynamic programming approaches Pang et al. (1981), Lagrangian relaxation methods Muckstadt and Koenig (1977); Bard (1988), and unit decommitment Tseng et al. (2000), among
An Improved MIP Formulation for the UC Problem
Atakan, S., Lulli, G., and Sen, S.

An Improved MIP Formulation for the UC Problem
Atakan, S., Lulli, G., and Sen, S.

others. For a detailed review, the reader is referred to Padhy (2004) and Saravanan et al. (2013). In recent years, mixed-integer programming (MIP) has become a popular tool for solving UC problems.

Nevertheless, it is well known that in general, mixed-integer programs are \( \mathcal{NP} \)-hard, and solving problems of realistic size, involving thousands of generators, remains challenging. Moreover, the introduction of variable energy resources (e.g., wind and solar) leads to circumstances in which predictions from deterministic models are subject to greater errors. In order to accommodate such changes, there have been recommendations that UC models should be solved using either the stochastic or the robust UC formulation (see, e.g., Cheung et al. (2015) and Bertsimas et al. (2013), respectively). In either case, the speed with which large scale UC models can be solved becomes important.

In this paper, we study the MIP formulations of the UC problem with components that are of general interest (e.g., essential operational constraints, piecewise-linear costs). Our study does not consider transmission networks. Within this context, the majority of the developments in the literature (e.g., Nowak and Rö misch (2000); Morales-España et al. (2013)) have been captured by the formulation appeared in Ostrowski et al. (2012). We will by and large consider their state-of-the-art formulation as a benchmark.

The contribution of this paper is twofold. We develop a new formulation that aims to obtain a tighter description of the UC polytope. Among the suggested improvements, the introduction of state-transition variables are of particular interest. First proposed by Pritsker et al. (1969) in the context of project scheduling, these binary variables, indexed with time periods \( t \), assume a value of 1 if an activity has been performed by time \( t \). In contrast, most formulations of the UC problem use decisions that reflect activities executed at time period \( t \). For instance, in many sequencing relations, it is important to know at any period of time if an activity has already been executed or not, regardless of its execution time. Moreover, the use of these by decision variables naturally introduces a graphical “sub-structure” into the problem formulation. In fact, the introduction of these variables was motivated by some successful applications in the field of transportation and scheduling (e.g., Bertsimas et al., 2014). In addition to the differences in the decision variables, the new formulation also introduces improved versions of the piecewise-linear production and start up costs.

Concomitant with the use of the new formulation, a certain set of constraints of the UC problem now becomes facet-defining by themselves. This, being our second contribution, leads to an important distinction from the state-of-the-art UC formulations. Such formulations are often improved by lifting the original constraints or appending additional valid inequalities. In contrast, in our proposed formulation certain facets of the UC polytope are naturally defined by the constraints of the problem.

Compared with the state-of-the-art formulation, the new formulation leads to shorter computational times, and the linear relaxations’ solutions are almost integral. This result provides empirical evidence that the polyhedral structure of the new formulation is tighter than the polyhedral structure of the state-of-the-art formulation. An important consequence of the properties of the new formulation is that it allows us to solve real-life instances in less than ten minutes on an ordinary desktop PC, compared to the state-of-the-art formulation, which is unable to provide a provably optimal solution within two hours of computing time on the same PC.

The rest of the paper is organized as follows. In §2, after presenting a prototypical formulation of UC, which resembles the first integer-programming (IP) formulation by Garver (1962), we present two formulations for the UC problem. The first one is the state-of-the-art formulation, which we refer to as the base formulation; whereas the second one, named new formulation, is the formulation that we propose. In §3, we report our computational experiences on both formulations by solving two classes of instances: (i) synthetic instances used by other academic researchers, and (ii) real-life instances made available by FERC for computational testing. Details on these two classes of instances, as well as our modifications, are given in the appendix. More specific conclusions of our study are provided in §4.

2. Alternative Mathematical Models for UC

Beginning with the IP formulation by Garver (1962), a plethora of mathematical models and formulations have been proposed in the literature to solve realistic instances of the UC problem. These formulations have
extended the work of Garver in several directions. The mathematical model has been enriched with many additional key aspects of the problem, thus making it more realistic. For instance, several operational and technological constraints have been included in the state-of-the-art UC models, and a certain effort has been spent in modeling operating costs. In addition to refining the mathematical modeling, a lot of the academic research in this field has been devoted to “strengthening” the corresponding formulations in order to achieve better computational performance.

With the purpose of providing a prototypical formulation of the UC problem, given a set $\mathcal{G}$ of generators, and a set $\mathcal{T}$ of time periods (which is commonly based on the hourly discretization of the planning horizon), we introduce the following sets of decision variables:

- $x_{g,t}$: State variable (1 if $g \in \mathcal{G}$ is operational at time $t \in \mathcal{T}$, 0 otherwise),
- $s_{g,t}$: Start up variable (1 if $g$ is turned on at time $t$, 0 otherwise),
- $z_{g,t}$: Shut down variable (if $g$ is turned off at time $t$, 0 otherwise),
- $p_{g,t}$: Amount of production in $g$ at time $t$.

The state variables are fundamental for the formulation of the operational constraints for the scheduling of generators. The start up and shut down variables are used to formulate the operating costs of the generators, whereas the production variables determine the dispatch amounts.

The objective of the mathematical model is to compute power production to meet the demand while minimizing the total costs. The total operational costs include both the production and the start up costs. Both of these costs are in general nonlinear. In our prototypical formulation, these costs are represented with the following (nonlinear) functions: $F_g(\cdot)$, for the start up costs, and $V_g(\cdot)$, for the production costs. The argument of the cost functions are respectively $i$- and $j$-dimensional vectors $x_{g,[t]} = (x_{g,-i}, \ldots, x_{g,t})$ and $p_{g,[t]} = (p_{g,-j}, \ldots, p_{g,t})$, for some $i \geq 0$ and $j \geq 0$, in order to emphasize the dependency of the costs respectively on the previous states and the production levels of the generator.

In the UC literature, other objectives have also been considered such as the minimization of no load or turn off costs, maximization of social welfare, or maximization of the profit of generator companies. However, these objective functions may require additional information. For instance, the latter two objectives require market information regarding the electricity prices (Shahidehpour et al., 2002, pages 5-8).

The prototypical formulation is given below:

$$\min \sum_{g \in \mathcal{G}} \sum_{t \in \mathcal{T}} F_g(x_{g,[t]}) + V_g(p_{g,[t]})$$

subject to:

1. $x_{g,t} - x_{g,t-1} = s_{g,t} - z_{g,t}$ \hspace{1cm} $\forall g \in \mathcal{G}, t \in \mathcal{T}$, (1a)
2. $p_{g,t} \geq \bar{C}_g x_{g,t}$ \hspace{1cm} $\forall g \in \mathcal{G}, t \in \mathcal{T}$, (1b)
3. $p_{g,t} \leq C_g x_{g,t}$ \hspace{1cm} $\forall g \in \mathcal{G}, t \in \mathcal{T}$, (1c)
4. $\sum_{g \in \mathcal{G}} p_{g,t} \geq d_t$ \hspace{1cm} $\forall t \in \mathcal{T}$, (1d)
5. $x_{g,t}, s_{g,t}, z_{g,t}, p_{g,t} \in \mathcal{P}_g \cap \mathcal{D}_t$ \hspace{1cm} $\forall g \in \mathcal{G}, t \in \mathcal{T}$, (1e)
6. $s_{g,t}, z_{g,t}, x_{g,t} \in \{0,1\}$ \hspace{1cm} $\forall g \in \mathcal{G}, t \in \mathcal{T}$, (1f)
7. $p_{g,t} \geq 0$ \hspace{1cm} $\forall g \in \mathcal{G}, t \in \mathcal{T}$, (1g)

where $d_t$ is the demand for electricity at time period $t$, $\bar{C}_g$ is the maximum generation capacity and $C_g$ is the minimum required production amount when the generator is operational.

The first set of constraints (1a), already introduced in the original formulation of Garver (1962), links the start up ($s_{g,t}$) and the shut down ($z_{g,t}$) variables to the state variables ($x_{g,t}$). Indeed, the start up and shut down variables are completely determined once the values of the state variables are fixed. They have been introduced exclusively to capture the transition of a generator from the idle state to the operational state (or vice versa), in order to formulate the operating costs of a generator. Constraint (1a) by itself does not forbid both the $s_{g,t}$
An Improved MIP Formulation for the UC Problem

Atakan, S., Lulli, G., and Sen, S.

and \(z_{g,t}\) variables to be simultaneously set to 1 whenever the generator remains operational (or idle). However, positive start up costs will ensure that an optimal solution would never lead to such assignments, as otherwise a start up cost would have been incurred due to \(s_{g,t}\) variables.

Constraints (1b) and (1c) model the lower and upper bounds, respectively. Constraints (1d) impose that the production level meets the demand for energy at each time period. Finally, constraints (1e) state that a feasible solution must belong to the polyhedra \(P_g\) and \(D_t\), which will be described in detail in the following paragraphs.

The careful reader may observe that the integrality requirements on the \(s_{g,t}\) and the \(z_{g,t}\) decision variables can be relaxed without invalidating the formulation. This observation was exploited by Carrión and Arroyo (2006) to formulate the UC problem with integrality restrictions on the state variables alone. However, the assumed benefit of using considerably smaller number of binary variables does not necessarily lead to superior computational performance, as evidenced by Ostrowski et al. (2012). This is because MIP solvers have become significantly more robust in recent years.

Before presenting the two formulations, we define the notation for some of the problem parameters:

- \(\bar{C}_g\): Maximum generation capacity,
- \(\bar{C}_g\): Minimum required production amount when the generator is operational,
- \(\bar{R}_g\): Ramp-up limit,
- \(\bar{R}_g\): Ramp-down limit,
- \(\bar{S}_g\): Ramp-up limit at the time when the generator is turned on (will be referred as start up limit),
- \(\bar{S}_g\): Ramp-down limit at the time when the generator is turned off (will be referred as shut down limit),
- \(UT_g\): Minimum required uptime before the generator can be turned off,
- \(DT_g\): Minimum required downtime before the generator can be turned on,

Note that the data of a typical UC instance usually obey the following relations: \(\bar{C}_g \geq \bar{R}_g \geq \bar{S}_g \geq C_g > 0\) and \(\bar{C}_g \geq R_g \geq S_g \geq C_g > 0\), \(\forall g \in G\). The condition \(C_g > 0\) avoids trivial solutions from the UC models where the generator is kept operational with no production.

### 2.1. The State-of-the-Art Formulation

The prototypical formulation in (1) has received considerable attention and has been the most prominent one over decades. Indeed, most of the formulations available in the literature have been to some extent derived from the prototypical formulation, and they differ from each other in the way they describe the two polyhedra \(D_g\) and \(P_g\). The former polyhedron models the reserve requirements, whereas the latter polyhedron captures all the operational and technological constraints.

In what follows, we will allow nonpositive indices of the decision variables to make it explicit that a solution of the problem might depend on the past states of the generators. We emphasize that in the actual implementation, we fix such variables to their realized values according to the available past data.

**Reserve requirements** A fundamental aspect of the UC problem is the modeling of reserve requirements (or operating reserves). These requirements are stand-by capacities that must be kept ready to generate energy to provide for unplanned outages of generating units. System operators use operating reserves to maintain system reliability and to ensure that the supply-demand balance is achieved seamlessly. Baldick (1995) and Nowak and Römisch (2000) provide examples on how to formulate these requirements.

In the prototypical model (1), the reserve requirements, which must be attained at each time period, are embodied in the polyhedron \(D_t\). To describe \(D_t\), we introduce the following set of variables:

- \(\bar{p}_{g,t}\): the maximum generation amount that \(g\) can supply at time \(t\),

and \(z_{g,t}\) variables to be simultaneously set to 1 whenever the generator remains operational (or idle). However, positive start up costs will ensure that an optimal solution would never lead to such assignments, as otherwise a start up cost would have been incurred due to \(s_{g,t}\) variables.

Constraints (1b) and (1c) model the lower and upper bounds, respectively. Constraints (1d) impose that the production level meets the demand for energy at each time period. Finally, constraints (1e) state that a feasible solution must belong to the polyhedra \(P_g\) and \(D_t\), which will be described in detail in the following paragraphs.

The careful reader may observe that the integrality requirements on the \(s_{g,t}\) and the \(z_{g,t}\) decision variables can be relaxed without invalidating the formulation. This observation was exploited by Carrión and Arroyo (2006) to formulate the UC problem with integrality restrictions on the state variables alone. However, the assumed benefit of using considerably smaller number of binary variables does not necessarily lead to superior computational performance, as evidenced by Ostrowski et al. (2012). This is because MIP solvers have become significantly more robust in recent years.

Before presenting the two formulations, we define the notation for some of the problem parameters:

- \(\bar{C}_g\): Maximum generation capacity,
- \(\bar{C}_g\): Minimum required production amount when the generator is operational,
- \(\bar{R}_g\): Ramp-up limit,
- \(\bar{R}_g\): Ramp-down limit,
- \(\bar{S}_g\): Ramp-up limit at the time when the generator is turned on (will be referred as start up limit),
- \(\bar{S}_g\): Ramp-down limit at the time when the generator is turned off (will be referred as shut down limit),
- \(UT_g\): Minimum required uptime before the generator can be turned off,
- \(DT_g\): Minimum required downtime before the generator can be turned on,

Note that the data of a typical UC instance usually obey the following relations: \(\bar{C}_g \geq \bar{R}_g \geq \bar{S}_g \geq C_g > 0\) and \(\bar{C}_g \geq R_g \geq S_g \geq C_g > 0\), \(\forall g \in G\). The condition \(C_g > 0\) avoids trivial solutions from the UC models where the generator is kept operational with no production.

### 2.1. The State-of-the-Art Formulation

The prototypical formulation in (1) has received considerable attention and has been the most prominent one over decades. Indeed, most of the formulations available in the literature have been to some extent derived from the prototypical formulation, and they differ from each other in the way they describe the two polyhedra \(D_g\) and \(P_g\). The former polyhedron models the reserve requirements, whereas the latter polyhedron captures all the operational and technological constraints.

In what follows, we will allow nonpositive indices of the decision variables to make it explicit that a solution of the problem might depend on the past states of the generators. We emphasize that in the actual implementation, we fix such variables to their realized values according to the available past data.

Reserve requirements A fundamental aspect of the UC problem is the modeling of reserve requirements (or operating reserves). These requirements are stand-by capacities that must be kept ready to generate energy to provide for unplanned outages of generating units. System operators use operating reserves to maintain system reliability and to ensure that the supply-demand balance is achieved seamlessly. Baldick (1995) and Nowak and Römisch (2000) provide examples on how to formulate these requirements.

In the prototypical model (1), the reserve requirements, which must be attained at each time period, are embodied in the polyhedron \(D_t\). To describe \(D_t\), we introduce the following set of variables:

- \(\bar{p}_{g,t}\): the maximum generation amount that \(g\) can supply at time \(t\),
In contrast to $p_{g,t}$, $\bar{p}_{g,t}$ does not represent the actual production but an achievable upper limit, i.e., the maximum production that generator $g$ can ramp up to at time $t$.

The $\bar{p}_{g,t}$ variables provide all the necessary information to model the reserve requirements. In a recent paper, Morales-Espa˜na et al. (2013) modeled the reserve requirements by means of the so-called reserve variables, which denote the portion of the generator capacity that is allocated for the reserve requirements (i.e., $\bar{p}_{g,t} - p_{g,t}$). However, our preliminary computational results indicated that the use of $\bar{p}_{g,t}$ variables provides slightly more consistent performance. In view of the definition of the $\bar{p}_{g,t}$ variables, the upper limits on the production levels are reformulated as follows:

\begin{align}
\bar{p}_{g,t} &\geq p_{g,t} \quad \forall g \in G, t \in T, \quad (2a) \\
\bar{p}_{g,t} &\leq \bar{C}_g x_{g,t} \quad \forall g \in G, t \in T. \quad (2b)
\end{align}

Denoting the amount of required reserve at time $t$ with $\rho_t$, the reserve requirements are enforced by the following constraint:

$$\sum_{g \in G} \bar{p}_{g,t} \geq d_t + \rho_t \quad \forall t \in T. \quad (3)$$

The polyhedron $D_t$ is the projection onto the space of the $x_{g,t}$, $z_{g,t}$, $s_{g,t}$, and $\bar{p}_{g,t}$ variables from the polyhedron defined by constraints (2)-(3).

**Operational and technological constraints** In the prototypical formulation given above, all the operational and technological constraints are captured in the polyhedra $P_g$. These constraints take into account the physical limitations of the generators that have to be satisfied in order to compute realistic production schedules. In this study, we focus on the fundamental version of the UC problem, which includes the minimum uptime/downtime and ramp up/down restrictions as its operational and technological constraints. Herein, we do not take into account more sophisticated constraints, such as constraints on transmission, power flow, line flow, and voltage limits; and/or total fuel and energy limit constraints. The interested reader may refer to Baldick (1995) for a general formulation containing these constraints.

The minimum uptime/downtime restrictions ensure that the on/off status of the generators do not change rapidly. Frequent state transitions have several adverse consequences including (i) increased operator stress, (ii) diminished generator life, and (iii) increased emission of pollutants during transient periods see (see Takriti and Birge, 2000). Such restrictions are quite practical and are included in many commercial tools. To model these restrictions, several formulations and valid inequalities were proposed in Takriti and Birge (2000); Arroyo and Conejo (2000); Lee et al. (2004). Following Rajan and Takriti (2005), the minimum uptime and downtime constraints are best modeled using the following constraints:

\begin{align}
\sum_{i=t-UT_g+1}^{t} s_{g,i} &\leq x_{g,t} \quad g \in G, \ t \in T, \quad (4a) \\
\sum_{i=t-DT_g+1}^{t} s_{g,i} &\leq 1 - x_{g,t-DT_g} \quad g \in G, \ t \in T. \quad (4b)
\end{align}

Observe that when the generator $g$ is operational at time $t$, the right-hand side of constraint (4a) is set to 1. In this case, the generator may have been turned on at most once in the last $UT_g$ time periods (due to minimum uptime restrictions). On the other hand, if it is idle at time $t$, it could not have been turned on in the last $UT_g$ time periods, as otherwise it should be operational at time $t$. Constraint (4b) is just a rewritten version of a similar constraint for the downtime requirements. It has been widely observed that these turn on/off inequalities significantly outperform other formulations (Ostrowski et al., 2012). Indeed, Rajan and Takriti (2005) showed that constraints (4) define facets of the minimum uptime/downtime polytope, that is, the polytope defined by the minimum uptime/downtime constraints.

The ramping restrictions limit the maximum increase/decrease in power attainable at any time instant $t$ and ensure that the generation requirements can be matched by the electricity production without exceeding the
generator limitations over extended periods of time. The ramping restrictions are expressed using the following constraints:

\[
\begin{align*}
    p_{g,t} - p_{g,t-1} &\leq S_g s_{g,t} + R_g x_{g,t-1} \quad \forall g \in \mathcal{G}, \ t \in \mathcal{T}, \\
    p_{g,t-1} - p_{g,t} &\leq S_g \bar{x}_{g,t} + R_g x_{g,t} \quad \forall g \in \mathcal{G}, \ t \in \mathcal{T}.
\end{align*}
\]  

Constraint (5a) ensures that generator \( g \) does not ramp up more than \( S_g \) if it has just been turned on, or \( R_g \) if it remains on at time \( t \). Similarly, constraint (5b) limits the decrease in the power output by \( R_g \) at any time that the generator remains operational. If the plant is turned off at \( t \) \((x_{g,t} = 0)\), then the output of the generator cannot be larger than \( S_g \) to obey the shut down limits. Observe that (5a) cannot be tight if the generator is turned off at time \( t \) \((p_{g,t} = 0)\) because \(-p_{g,t-1} \leq -C_g < 0\), but the right-hand-side is \( R_g > 0 \). A similar argument also applies to (5b) when the generator is turned on at time \( t \). In fact, these (along with many others) were the motivation behind studying tighter ramping constraints and valid inequalities in Ostrowski et al. (2012); Jiang et al. (2014). In particular, Damcı-Kurt et al. (2014) provided the strengthened ramp up/down constraints

\[
\begin{align*}
    p_{g,t} - p_{g,t-1} &\leq (S_g - R_g - C_g)s_{g,t} + (R_g + C_g)x_{g,t} - C_g x_{g,t-1} \quad \forall g \in \mathcal{G}, \ t \in \mathcal{T}, \\
    p_{g,t-1} - p_{g,t} &\leq (S_g - R_g - C_g)\bar{x}_{g,t} + (R_g + C_g)x_{g,t} - C_g x_{g,t-1} \quad \forall g \in \mathcal{G}, \ t \in \mathcal{T},
\end{align*}
\]

which were proved to be facet-defining for the two-period ramp-up and ramp-down polytopes (the polytopes of the UC problem that are limited to two consecutive time periods and consider only the ramp-up and ramp-down constraints, respectively).

In case the model includes reserve requirements, the ramping constraints have to be modified in order to guarantee that the generators can actually contribute to the operating reserve. More specifically, the left-hand side term of the ramp-up constraints (5a) and (6a), which account for the variation of electricity produced at two consecutive time periods (i.e., \( p_{g,t} - p_{g,t-1} \)), has to be replaced with the following difference: \( \bar{p}_{g,t} - \bar{p}_{g,t-1} \).

**Linearization of the objective function** The nonlinear functions in the objective are usually approximated using piecewise-linear convex functions. Here, we present the scheme described in Nowak and Römisch (2000), which requires the following notation:

- \( f^\tau_g \): Start up cost when the generator has been idle for \( \tau \) time units,
- \( r^\text{max}_g \): Index of the maximum start up cost,
- \( v^\kappa_{g} \): Unit cost of generation at the \( \kappa \)th level of production,
- \( k^\text{max}_g \): Index of the maximum variable cost,
- \( V_g(\cdot) \): Cumulative cost of generation for a given production amount (excludes start up costs),
- \( p^\kappa_g \): Maximum amount of production that is achievable with a unit cost of at most \( v^\kappa_{g} \).

The cost coefficients are assumed to obey the relations \( f^\tau_g \geq \cdots \geq f^1_g \) and \( v^{\text{max}}_g \geq \cdots \geq v^1_g \), \( \forall g \in \mathcal{G}, \) indicating that the start up costs tend to increase as the idle time of the generators grows, and the unit cost of production increases as the production amounts approach the generator capacities.

The start up cost incurred by a generator \( g \) at time \( t \) is accounted by the non-negative variable \( f_{g,t} \), which satisfies the following constraints:

\[
    f_{g,t} \geq f^\tau_g \left( x_{g,t} - \sum_{i=1}^\tau x_{g,t-i} \right) \quad \forall g \in \mathcal{G}, \ t \in \mathcal{T}, \text{ and } \forall \tau = \{1, \ldots, r^\text{max}_g\}.
\]  

Observe that it is sufficient to write (7) only for those \( \tau \in \{1, \ldots, r^\text{max}_g\} \) when \( f^\tau_g \neq f^{\tau-1}_g \).

The production cost at time \( t \) is considered to be a function of \( p_{g,t} \), and is again approximated using a piecewise-linear convex function. The production levels \( p^\kappa_g \) (with \( \kappa = \{1, \ldots, k^\text{max}_g\} \)) are the breakpoints where
the slope of the piecewise-linear function changes in $v^\kappa_g$. Therefore, the production costs of generator $g$ at time $t$ are accounted by non-negative variable $v_{g,t}$, which satisfies the following constraints:

$$v_{g,t} \geq v^\kappa_g(p_{g,t} - p^\kappa_{g}^{-1}) + \hat{V}_g(p^\kappa_{g}^{-1}) \quad \forall g \in \mathcal{G}, \ t \in \mathcal{T}, \text{ and } \forall \kappa = \{1, \ldots, \kappa_g^{\text{max}}\}. \tag{8}$$

We assume that $p^0_{g} = 0$ (and hence, $\hat{V}_g(p^0_{g}) = 0$).

The base formulation (Ostrowski et al., 2012) The description of the state-of-the-art formulation is summarized below.

$$\min \sum_{t \in \mathcal{T}} \sum_{g \in \mathcal{G}} f_{g,t} + v_{g,t}$$

subject to:

- (1a) state constraints,
- (1b) capacity constraints,
- (1d), (2) demand constraints and reserve requirements,
- (4) min up/downtime constraints,
- (5), (7) ramp up/down constraints,
- (8) piecewise-linear start up and production cost functions,

$x_{g,t}$, $s_{g,t}$, $z_{g,t} \in \{0, 1\}, \forall g \in \mathcal{G}, \ t \in \mathcal{T}$, integrality restrictions

$p_{g,t}$, $\bar{p}_{g,t}$, $f_{g,t}$, $v_{g,t} \geq 0, \forall g \in \mathcal{G}, \ t \in \mathcal{T}$ nonnegativity restrictions.

2.2. The New Formulation

In this section, we present a new formulation for the UC problem that is primarily based on the introduction of new variables. More specifically, we use different variables when representing the state of a generator and its production level. To better highlight the differences and similarities with the state-of-the-art formulation, in what follows, we discuss each component separately.

State-transition variables We define the following set of decision variables:

$$\tilde{x}_{g,t}: \text{ 1 if } g \in \mathcal{G} \text{ remains operational at time } t \in \mathcal{T}, \text{ 0 otherwise.}$$

Note that $\tilde{x}_{g,t}$, along with $s_{g,t}$ and $z_{g,t}$ do not capture the state of a generator, but rather capture the state-transition between two consecutive time periods. We illustrate this in Figure 1 where the possible states of a generator are depicted by nodes, and the feasible state-transitions are depicted by arcs. The transition in which the generator remains idle (off), is marked with a dashed line because no decision variable is attached to it. If defined, the decision variable of this transition would be completely determined by the values of the other variables. In fact, exactly one of the state transitions occurs at each time period $t$ and the corresponding decision variable is set to 1. Hence, if none of the described variables are set to 1, the transition corresponding to the dashed line is said to occur. From this description, the following lemma trivially follows.

Lemma 1. For each generator $g \in \mathcal{G}$ and time period $t \in \mathcal{T}$, the following inequality is valid,

$$s_{g,t} + \tilde{x}_{g,t} + z_{g,t} \leq 1.$$
operational at time $t-1$ then it must either be turned off or remain on at time $t$. The constraints (9) will be referred to as state-transition constraints.

The linear map $(x_{g,t}) \rightarrow (s_{g,t} + \bar{x}_{g,t})$ allows us to translate any constraints derived for the base formulation into the constraints of the new formulation. Although the essence of the constraints are different, it is easy to verify that (9) can also be derived from (1a) by the transformation described above.

**Production-related variables** In order to model the reserve requirements, we again make use of the $\bar{p}_{g,t}$ variables, which were introduced earlier for the base formulation. In the new formulation, we also make use of the following variables:

$$ r_{g,t}: \text{production amount beyond } \bar{C}_g \text{ of generator } g \in \mathcal{G} \text{ at time } t \in \mathcal{T}. $$

The above variable, now only account for the “variable” amount of power output, that is, the amount produced beyond $\bar{C}_g$. In fact, due to the minimum production requirements, any operational generator $g$ produces an output of at least $\bar{C}_g$, an amount of power that can be associated with the variables $s_{g,t}$ and $\bar{x}_{g,t}$. Using these production variables, the minimum production constraints are immediately satisfied, and therefore becomes redundant. This allows us to remove $|\mathcal{G}| \times |\mathcal{T}|$ constraints from the formulation without sacrificing its fidelity. In view of the decision variables defined above, the constraints that model the demand requirements and generation limits can be reformulated as follows:

$$ \bar{p}_{g,t} \geq r_{g,t} + \bar{C}_g(s_{g,t} + \bar{x}_{g,t}) \quad \forall g \in \mathcal{G}, \; t \in \mathcal{T}. \quad (10a) $$

$$ \bar{p}_{g,t} \leq \bar{S}_g(s_{g,t} + \bar{x}_{g,t}) \quad \forall g \in \mathcal{G}, \; t \in \mathcal{T}. \quad (10b) $$

$$ \sum_{g \in \mathcal{G}} r_{g,t} + \bar{C}_g(s_{g,t} + \bar{x}_{g,t}) \geq Dem_t \quad \forall t \in \mathcal{T}. \quad (11) $$

Constraints (10) are the generation limits, while constraints (11) are the demand constraints. As a tighter alternative, since $\bar{S}_g \leq \bar{C}_g$, the following constraint

$$ \bar{p}_{g,t} \leq \bar{S}_g s_{g,t} + \bar{C}_g \bar{x}_{g,t} \quad \forall g \in \mathcal{G}, \; t \in \mathcal{T}. \quad (12) $$

can be used in lieu of (10b). Indeed (12) is equivalent to constraints (5) of Damci-Kurt et al. (2014), where they were proved to be the facets of the two-period ramping polytope. Finally, the reserve requirement constraints (3) of the base formulation are kept the same in the new formulation.

Before proceeding, we should emphasize that the idea to treat the minimum required generation and the “variable” generation amounts separately was originally considered in Garver (1962). Nevertheless, over the years this idea has been set aside, until the recent study of Morales-España et al. (2013). In that paper, the authors also consider $r_{g,t}$ to denote the production amounts above $\bar{C}_g$, and model the capacity and demand constraints accordingly. However these variables are not exploited further. As we will see shortly, by analyzing
the ramping constraints and the piecewise-linear variable costs, one can achieve improved descriptions of the UC problem.

**Operational and technological constraints** We now formulate the minimum uptime/downtime and the ramping restrictions. In order to model the former, we translate inequalities (4) given for the base formulation as presented below:

\[
\begin{align*}
\sum_{i=t-UT_g+1}^{t-1} s_{g,i} & \leq \tilde{x}_{g,t} \quad g \in \mathcal{G}, \ t \in \mathcal{T}, \\
\sum_{i=t-DT_g}^{t} s_{g,i} & \leq 1 - \tilde{x}_{g,t-1} \quad g \in \mathcal{G}, \ t \in \mathcal{T}.
\end{align*}
\] (13a)

Constraint (13a) ensures that if the generator remains on, it could have been turned on at most once in the previous \(UT_g - 1\) time periods. If it does not “remain-on”, then it could not have been turned on in these time periods due to minimum uptime restrictions. Similarly, constraint (13b) ensures that if the generator remains on, it cannot be restarted in the current time period or in the next \(DT_g\) time periods due to minimum downtime restrictions. On the other hand, if it does not “remain-on”, it can be turned on at most once in these time periods, due to minimum uptime and downtime restrictions.

In order to model the ramping restrictions, we take advantage of both the state-transition and the production variables, and provide the following constraints:

\[
\begin{align*}
p_{g,t} - r_{g,t-1} & \leq \bar{S}_g s_{g,t} + (\bar{R}_g + C_g) \tilde{x}_{g,t} \quad \forall g \in \mathcal{G}, \ t \in \mathcal{T}.
\end{align*}
\] (14a)

The coefficient of the “remain-on” variable is increased by \(\bar{C}_g\) because the \(r_{g,t-1}\) variable only accounts for the power generation beyond \(C_g\). When a generator is turned on in period \(t\), it should have been idle in period \(t-1\). Therefore \(r_{g,t-1} = 0\) and no increment for the \(s_{g,t}\) coefficient is necessary. The following propositions and the ensuing remark are provided to support the strength of the proposed ramping inequalities.

**Proposition 1.** For each \(g \in \mathcal{G}, \ t \in \mathcal{T}\), the ramp-up inequality (14a) is valid and dominates constraint

\[
p_{g,t} - p_{g,t-1} \leq \bar{S}_g s_{g,t} + \bar{R}_g \tilde{x}_{g,t}
\] (P1a)

which is equivalent to (5a) when the variables of the new formulation are used.

**Proof.** It is straightforward to verify the validity of constraints (14a). The inequality limits the increase in production by \(\bar{S}_g\) if the generator has just been turned on (in this case \(r_{g,t-1} = 0\), or by \((\bar{R}_g + C_g)\) if it remains on. As also pointed out in §2.1, the ramp-up inequality (P1a) is inactive if the generator has just been turned off; but it can be strengthened by lifting it to the space of \(z_{g,t}\) variables.

\[
p_{g,t} - [r_{g,t-1} + C_g(s_{g,t-1} + \tilde{x}_{g,t-1})] \leq \bar{S}_g s_{g,t} + \bar{R}_g \tilde{x}_{g,t} - C_g \tilde{z}_{g,t}.
\] (P1b)

The term within brackets is \(p_{g,t-1}\). To verify that (P1b) is valid, it is sufficient to consider the case where \(z_{g,t} = 1\) (the case \(z_{g,t} = 0\) is trivial). As the generator is not operational at time period \(t\), \(p_{g,t}\) is 0 and the inequality reduces to \(r_{g,t-1} + C_g(s_{g,t-1} + \tilde{x}_{g,t-1}) = p_{g}(t-1) \geq C_g\), which is the minimum production requirement. Therefore (P1b) is valid and dominates (P1a), which is easy to see as the left-hand side of (P1a) assumes no smaller value than the left-hand side of (P1b). Constraints (14a) can then be obtained from (P1a) by simple algebra:

\[
\begin{align*}
p_{g,t} - r_{g,t-1} & \leq \bar{S}_g s_{g,t} + \bar{R}_g \tilde{x}_{g,t} - C_g z_{g,t} + C_g(s_{g,t-1} + \tilde{x}_{g,t-1}) \quad (P1c) \\
& = \bar{S}_g s_{g,t} + \bar{R}_g \tilde{x}_{g,t} - C_g z_{g,t} + C_g(z_{g,t} + \tilde{x}_{g,t}) \quad (P1d) \\
& = \bar{S}_g s_{g,t} + (\bar{R}_g + C_g) \tilde{x}_{g,t}.
\end{align*}
\]
A similar analysis can also be performed for the ramp-down constraints, which is given in the following proposition.

**Proposition 2.** For each \( g \in \mathcal{G} \), \( t \in \mathcal{T} \), the ramp-down inequality

\[
r_{g,t-1} - r_{g,t} \leq (S_g - C_g)z_{g,t} + B_g \tilde{x}_{g,t}\]

is valid and dominates the seed inequality \( p_{g,t-1} - p_{g,t} \leq S_g z_{g,t} + B_g \tilde{x}_{g,t} \).

**Proof.** Consider the seed inequality:

\[
[r_{g,t-1} + C_g(s_{g,t-1} + \tilde{x}_{g,t-1})] - [r_{g,t} + C_g(s_{g,t} + \tilde{x}_{g,t})] \leq S_g z_{g,t} + B_g \tilde{x}_{g,t},
\]

which can be strengthened by lifting it to the space of the \( s_{g,t} \) variables,

\[
[r_{g,t-1} + C_g(s_{g,t-1} + \tilde{x}_{g,t-1})] - [r_{g,t} + C_g(s_{g,t} + \tilde{x}_{g,t})] \leq S_g z_{g,t} + B_g \tilde{x}_{g,t} - C_g s_{g,t}.
\]

When \( s_{g,t} = 1 \), the power output at time \( t - 1 \) must be 0, and the first term in brackets disappears. Since \( \tilde{x}_{g,t} = z_{g,t} = 0 \), the inequality simply reduces to \( r_{g,t} \geq 0 \), which confirms its validity. By simple algebra, we obtain

\[
r_{g,t-1} - r_{g,t} \leq S_g z_{g,t} + (B_g + C_g) \tilde{x}_{g,t} - C_g s_{g,t-1} + \tilde{x}_{g,t-1}
\]

\[
= S_g z_{g,t} + (B_g + C_g) \tilde{x}_{g,t} - C_g (z_{g,t} + \tilde{x}_{g,t})\]

\[
= (S_g - C_g) z_{g,t} + B_g \tilde{x}_{g,t}\]

where \((P2)\) is achieved using \((9)\). \(\blacksquare\)

Using the relations \((1a)\) and \((9)\), it is trivial to verify the following remark.

**Remark 1.** The ramping inequalities \((14a)\) and \((14b)\) of the new formulation are equivalent to the facet-defining two-period ramping inequalities of the base formulation, proposed by Damcı-Kurt et al. (2014), i.e., constraints \((6a)\) and \((6b)\).

**Linearization of the objective function** We now model the piecewise-linear approximation of the start up and the production cost functions, following the same approach described for the base formulation. We begin with the start up costs. Observe that, when turned on, the start up cost of a generator will at least be its warm-start cost, i.e., the start up cost incurred when the generator has not cooled down since the previous operational state. Due to the minimum downtime restriction, this cost is at least \( f c_g^{DT_g} \). Furthermore, it can be associated with the start up variable and accounted directly in the objective function using the additional term \( f c_g^{DT_g} \cdot s_{g,t} \). Therefore, the variable \( f_{g,t} \geq 0 \) will now represent the extra cost of turning on a generator which has been idle for some time that is longer than its minimum downtime \((DT_g)\). In view of this observation, variables \( f_{g,t} \geq 0 \) now satisfy the following constraints:

\[
f_{g,t} \geq (f c_g^{DT_g} - f c_g^{DT_g}) \left( s_{g,t} - \sum_{i=DT_g}^\tau s_{g,t-i} - \tilde{x}_{g,t-i} \right) \quad \forall g \in \mathcal{G}, \ t \in \mathcal{T}, \text{ and } \tau \in \{DT_g, \ldots, \tau_{max}\}.
\]

We note that, in general, the number of constraints to model the piecewise-linear cost functions can make mixed-integer programs extremely large. The representation above mitigates this potential issue by using \( |\mathcal{T}| \times \sum_{g \in \mathcal{G}} DT_g \) less number of constraints.

A similar observation is also available for the production costs. Recall that the total production of a generator is now accounted by two terms, \( r_{g,t} \) and \( C_g(s_{g,t} + \tilde{x}_{g,t}) \). As long as the generator is operational, the cost of producing the initial \( C_g \) units of energy is fixed and can be computed a-priori (herein denoted as \( V_g(C_g) \)). This cost can be associated with the \( s_{g,t} \) and \( \tilde{x}_{g,t} \) variables and directly accounted in the objective function.
Therefore, the variables \( v_{g,t} \) now provides the cost of producing an amount of energy in addition to \( C_g \). This cost is computed by the following set of constraints:

\[
v_{g,t} \geq v_{g}^{\kappa} (r_{g,t} + C_g - p_g^{\kappa-1}) + \bar{V}_g (p_g^{\kappa-1}) - \bar{V}_g (C_g) \quad \forall g \in \mathcal{G}, t \in \mathcal{T}, \text{ and } \forall \kappa = \{l_g, \ldots, \kappa_g^{\max}\}. \tag{16}
\]

Above, \( l_g \) \((\geq 1)\) is the value of the index \( \kappa \) for which \( p_g^{l_g-1} = C_g \). For values of \( \kappa \in \{1, \ldots, l_g - 1\} \), constraints (16) are omitted because their right-hand side will be no larger than zero and will be trivially satisfied. The results in §3 show that a significant portion of these constraints will be redundant in real-life instances as some of the given cost coefficients correspond to generation amounts less than \( C_g \).

**The new formulation** A summary of the new formulation is provided below:

\[
\begin{align*}
\min & \quad \sum_{t \in \mathcal{T}} \sum_{g \in \mathcal{G}} f_{g,t} + v_{g,t} + (f_{g}^{DT_t} + C_g) s_{g,t} + C_g \tilde{x}_{g,t} \\
\text{subject to:} & \quad \text{(9) state transition constraints,} \\
& \quad \text{(10) capacity constraints,} \\
& \quad \text{(3), (11) reserve requirements and demand constraints,} \\
& \quad \text{(13) min up/downtime constraints,} \\
& \quad \text{(14) ramp up/down constraints,} \\
& \quad \text{(15), (16) piecewise-linear start up and production cost functions,} \\
& \quad \bar{x}_{g,t}, s_{g,t}, z_{g,t} \in \{0, 1\}, \quad \forall g \in \mathcal{G}, t \in \mathcal{T}, \\
& \quad r_{g,t}, \tilde{p}_{g,t}, f_{g,t}, v_{g,t} \geq 0, \quad \forall g \in \mathcal{G}, t \in \mathcal{T}, \\
& \quad \text{integrality restrictions,} \\
& \quad \text{nonnegativity restrictions.}
\end{align*}
\]

Observe that the new formulation does not include the constraints \( s_{g,t} + \bar{x}_{g,t} + z_{g,t} \leq 1, \quad \forall g \in \mathcal{G}, t \in \mathcal{T} \) because they are already implied, as formalized in the following proposition.

**Proposition 3.** The constraints

\[
s_{g,t} + \bar{x}_{g,t} + z_{g,t} \leq 1 \quad \forall g \in \mathcal{G}, t \in \mathcal{T},
\]

are implied by the minimum downtime constraints (13b) and the state-transition constraints (9).

**Proof.** Consider constraint (13b) for the time period index \( t + DT_g - 1 \),

\[
1 \geq \tilde{x}_{g,t-1} + \sum_{i=t-1}^{t+DT_g-1} s_{g,i} = \tilde{x}_{g,t-1} + s_{g,t-1} + \sum_{i=t}^{t+DT_g-1} s_{g,i} = [\text{by (9)}] \\
= z_{g,t} + \bar{x}_{g,t} + s_{g,t} + \sum_{i=t+1}^{t+DT_g-1} s_{g,i} \geq z_{g,t} + \bar{x}_{g,t} + s_{g,t}.
\]

\[
\blacksquare
\]

**Remark 2.** If \( DT_g = 1 \), constraints \( s_{g,t} + \bar{x}_{g,t} + z_{g,t} \leq 1, \forall t \in \mathcal{T} \), are equivalent to the minimum downtime constraints of generator \( g \).

Finally, observe that the use of the new decision variables does not affect the polyhedral properties of the minimum uptime/downtime constraints (13) and the ramping constraints (14), as has been studied for the state-of-the-art formulation. Indeed the following results hold true.

**Proposition 4.** The minimum up/downtime constraints (13) are the facets of the minimum up/downtime polytope.

**Proof.** See Appendix A.1.

Following the definition in Damci-Kurt et al. (2014) for the two-period ramping polytope of generator \( g \), as the convex hull of the ramping constraints and the production limits are defined onto the space of \( \tilde{p}_{g,t}, r_{g,t-1}, r_{g,t}, s_{g,t}, \bar{x}_{g,t} \) and \( z_{g,t} \) decision variables, the following proposition holds true.
Proposition 5. The ramping inequalities, (14), and the the generation limit constraints, (10a) and (12), are the facets of the two-period ramping polytope.

Proof. See Appendix A.2. ■

3. Computational Experiments

Our analysis will focus on two classes of instances, where the former is composed of synthetic instances and the latter includes real-life instances. The class of synthetic instances comprises, among others, the 20 instances of Ostrowski et al. (2012), which are based on Carrión and Arroyo (2006). These instances refer to daily UC problems, i.e., instances in which the time horizon is restricted to one day, and are characterized by increasing numbers of generators. We have opted to use these instances as they have also been tested in other recent studies. In order to test the scalability of our model, we also extended the instances of Ostrowski et al. to week-long time horizon. Note that, instances with a longer time horizon are computationally more challenging not simply because of the increased sizes of the formulations but also due to the daily trends and fluctuations in the demand. The synthetic instances involve significant amounts of symmetry in the problem parameters, which is known to deteriorate the performance of branch-and-bound based algorithms. Moreover, to perform statistically meaningful comparisons, we have generated 100 synthetic instances of the largest size, i.e., the week-long instance with 187 generators. The procedures followed to generate synthetic instances are detailed in Appendix B.1. For the original data generation procedure, the reader is referred to the original papers. The real-life instances include nearly a thousand generators and the data have been obtained through the FERC. The details of the original instance can be found in Krall et al. (2012), and the modifications and extensions that we have made are listed in Appendix B.2.

Both formulations require the same number of (binary) variables, however the numbers of constraints are different. In particular, the new formulation contains a smaller number of constraints, on average 24% less, in both the day-long and the week-long synthetic instances. For the real-life instances, we give statistics of each formulation in Table 1. As the table demonstrates, a real-life UC problem can be quite demanding as the number of variables and constraints can grow to be huge.

| Formulation | $|T|$ | Variables | Binaries | Constraints |
|-------------|-----|-----------|----------|-------------|
| Base        | 24  | 157,753   | 67,608   | 287,108     |
|             | 168 | 1,104,265 | 473,256  | 2,016,548   |
| New         | 24  | 157,753   | 67,608   | 225,188     |
|             | 168 | 1,104,265 | 473,256  | 1,583,108   |

Table 1: The number of (binary) variables and constraints of the real-life instances.

All runs were performed on a single thread of a Dell Desktop PC with Intel® Core™ i7-3770S CPU @ 3.10 GHz, 7.68 GB of RAM, and running Ubuntu Linux 12.04.3 LTS. The mixed-integer programs were solved through the Concert Technology class library of IBM ILOG CPLEX 12.5.1 (CPLEX). The default parameters of CPLEX were preserved.

We first provide an empirical evaluation of the base formulation chosen for this computational study. The base formulation, given at the end of §2.1, does not use the two-period ramping inequalities (6) because, at least within the instances we have tested, they do not provide a consistent advantage with respect to the ramping inequalities (5). We begin with comparing these two versions of the base formulations, i.e., one using (5) and the other using (6), by solving the 100 week-long instances with 187 generators. Figure 2 plots the percentage differences of the computational times required to solve each of the 100 instances. Letting $T_B$ and $T_{B_2}$ denote the computational times with (5) and (6), respectively, these percentage differences are computed as $100 \times (T_B - T_{B_2})/T_{B_2}$. Positive (resp. negative) values of this ratio denote an improvement (resp. degradation) on the computational performance obtained by using (6). In several instances, while the use of (6) significantly improves the computational performance, it is also evident that in more than half of the tested instances, they did not enhance the performance. This is highlighted by the box plot displayed in Figure 2 (right). The average
values of $T_B$ and $T_{B_2}$ are respectively 196.3 and 231.6 seconds. This behavior can be partially ascribed to the weakened polyhedral properties of constraints (6) when incorporated into a UC formulation with additional components (such as the modeling of the reserve requirements).

![Figure 2: Effect of the two-period ramp-up and ramp-down inequalities on the base formulation.](image)

<table>
<thead>
<tr>
<th>Instance</th>
<th># Generators</th>
<th>Cut Generation</th>
<th>Nodes</th>
<th>Time (s)</th>
<th>Cut Generation</th>
<th>Nodes</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily (</td>
<td>T</td>
<td>= 24)</td>
<td></td>
<td>Base</td>
<td>New</td>
<td>Base</td>
<td>New</td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>61</td>
<td>1</td>
<td>0.7</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>44</td>
<td>532</td>
<td>0</td>
<td>0.8</td>
<td>47</td>
<td>0</td>
<td>0.4</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>328</td>
<td>1.0</td>
<td>0</td>
<td>2,593</td>
<td>491</td>
<td>41.3</td>
</tr>
<tr>
<td>5</td>
<td>49</td>
<td>844</td>
<td>1.7</td>
<td>16</td>
<td>2,619</td>
<td>140</td>
<td>23.1</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>910</td>
<td>2.4</td>
<td>99</td>
<td>3,023</td>
<td>523</td>
<td>64.2</td>
</tr>
<tr>
<td>7</td>
<td>51</td>
<td>269</td>
<td>0.7</td>
<td>0</td>
<td>1,052</td>
<td>121</td>
<td>83</td>
</tr>
<tr>
<td>8</td>
<td>51</td>
<td>739</td>
<td>0.9</td>
<td>0</td>
<td>1,515</td>
<td>9.7</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>52</td>
<td>506</td>
<td>0.8</td>
<td>0</td>
<td>1,760</td>
<td>15.9</td>
<td>108</td>
</tr>
<tr>
<td>10</td>
<td>54</td>
<td>815</td>
<td>1.2</td>
<td>5</td>
<td>2,126</td>
<td>16.4</td>
<td>129</td>
</tr>
<tr>
<td>11</td>
<td>132</td>
<td>311</td>
<td>2.3</td>
<td>0</td>
<td>2,798</td>
<td>54.0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>156</td>
<td>1,027</td>
<td>3.2</td>
<td>0</td>
<td>2,858</td>
<td>59.1</td>
<td>93</td>
</tr>
<tr>
<td>13</td>
<td>156</td>
<td>1,729</td>
<td>3.9</td>
<td>0</td>
<td>2,104</td>
<td>45.2</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>165</td>
<td>3,372</td>
<td>8.0</td>
<td>205</td>
<td>6,883</td>
<td>109.2</td>
<td>243</td>
</tr>
<tr>
<td>15</td>
<td>167</td>
<td>2,350</td>
<td>5.1</td>
<td>3</td>
<td>8,206</td>
<td>124.5</td>
<td>265</td>
</tr>
<tr>
<td>16</td>
<td>172</td>
<td>1,552</td>
<td>4.8</td>
<td>0</td>
<td>2,652</td>
<td>30</td>
<td>114.0</td>
</tr>
<tr>
<td>17</td>
<td>182</td>
<td>2,882</td>
<td>9.9</td>
<td>15</td>
<td>8,120</td>
<td>116.6</td>
<td>180</td>
</tr>
<tr>
<td>18</td>
<td>182</td>
<td>2,276</td>
<td>7.1</td>
<td>0</td>
<td>10,630</td>
<td>201.6</td>
<td>396</td>
</tr>
<tr>
<td>19</td>
<td>183</td>
<td>1,918</td>
<td>7.1</td>
<td>0</td>
<td>8,381</td>
<td>123.2</td>
<td>341</td>
</tr>
<tr>
<td>20</td>
<td>187</td>
<td>3,120</td>
<td>9.7</td>
<td>157</td>
<td>9,764</td>
<td>102.5</td>
<td>393</td>
</tr>
<tr>
<td>Average:</td>
<td>1,277.1</td>
<td>0.8</td>
<td>3.6</td>
<td>27.4</td>
<td>3.8</td>
<td>1.2</td>
<td>3,974.7</td>
</tr>
<tr>
<td># Nonzeros:</td>
<td>19</td>
<td>2</td>
<td>-</td>
<td>8</td>
<td>3</td>
<td>-</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 2: The number of cutting planes generated, nodes explored, and the time elapsed for CPLEX to reach optimality with the base, and new formulations.

Now, we compare and contrast the computational performance of the base and the new formulation presented in §2 by solving the instances of Ostrowski et al. (2012) and their weekly counterparts. Table 2 reports the total number of cuts and branch-and-bound (B&B) nodes generated by the CPLEX branch-and-bound algorithm, and the elapsed time until optimality was established. With the exception of the smallest instances, i.e., the ones with 28 generators, the new formulation clearly outperforms the base formulation in terms of the computational time. The larger the instance is, the more marked the improvement becomes in terms the time expended, as well as the number of B&B nodes explored. This is particularly evident in the week-long instances. Using the new
An Improved MIP Formulation for the UC Problem

Atakan, S., Lulli, G., and Sen, S.

formulation, almost all of the synthetic instances were solved within a half minute, whereas the solution time of the base formulation may even exceed 200 seconds. Moreover, we observe that the number of cuts generated for the base formulation are higher than the number of cuts generated for the new formulation. For the week-long instances, we have detailed the numbers of cutting planes of each type in Table 3. More specifically, Table 3 displays the following types of cuts: implied bound (IB), flow cover, mixed-integer rounding (MIR) and Gomory. Other types of cuts have not been listed because they only account for 0.2% of all generated cuts. As far as the numbers of B&B nodes are concerned, the new formulation achieves a substantial reduction in the week-long instances. These results unveil the potential of our formulation to scale up to even larger instances that are solved by the power industry.

<table>
<thead>
<tr>
<th>Implied Bound</th>
<th>Flow Cover</th>
<th>MIR</th>
<th>Gomory</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>161</td>
<td>2</td>
<td>99</td>
</tr>
<tr>
<td>2</td>
<td>690</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>896</td>
<td>101</td>
<td>89</td>
</tr>
<tr>
<td>4</td>
<td>1,904</td>
<td>60</td>
<td>97</td>
</tr>
<tr>
<td>5</td>
<td>2,013</td>
<td>51</td>
<td>97</td>
</tr>
<tr>
<td>6</td>
<td>2,018</td>
<td>76</td>
<td>96</td>
</tr>
<tr>
<td>7</td>
<td>640</td>
<td>15</td>
<td>98</td>
</tr>
<tr>
<td>8</td>
<td>1,133</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>1,372</td>
<td>43</td>
<td>97</td>
</tr>
<tr>
<td>10</td>
<td>1,397</td>
<td>40</td>
<td>97</td>
</tr>
<tr>
<td>11</td>
<td>2,193</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>1,932</td>
<td>29</td>
<td>98</td>
</tr>
<tr>
<td>13</td>
<td>1,420</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>14</td>
<td>5,147</td>
<td>141</td>
<td>97</td>
</tr>
<tr>
<td>15</td>
<td>6,172</td>
<td>152</td>
<td>98</td>
</tr>
<tr>
<td>16</td>
<td>1,687</td>
<td>440</td>
<td>74</td>
</tr>
<tr>
<td>17</td>
<td>6,440</td>
<td>114</td>
<td>98</td>
</tr>
<tr>
<td>18</td>
<td>8,177</td>
<td>241</td>
<td>97</td>
</tr>
<tr>
<td>19</td>
<td>6,486</td>
<td>195</td>
<td>97</td>
</tr>
<tr>
<td>20</td>
<td>7,572</td>
<td>243</td>
<td>97</td>
</tr>
</tbody>
</table>

Avg: | - | - | 96.3 | - | - | 75.5 | - | - | 95.4 | - | - | 74.1 |

Table 3: Number of cutting planes generated by CPLEX in each major type and the percentage of reduction achieved by using the new formulation (|T| = 168).

Before proceeding, we should clarify the discrepancy in the performance of the base formulation between the results reported in Table 2 and the results presented in Ostrowski et al. (2012) for |T| = 24. In particular, the performance of the base formulation reported in that paper are inferior to those that we have observed in our experiments. The main reason for this disparity is that our implementation incorporates the initial states of the generators into the models. In contrast, Ostrowski et al. (2012) assume that generators can be turned on and off at the beginning of the time horizon regardless of their past statuses. This, as confirmed by our preliminary studies, leads to computationally more demanding models. We note that in our study, we consider the initial states of the generators, as provided in Carrión and Arroyo (2006).

To provide an empirical evaluation on the strength of the polyhedral approximation of the new formulation, we present an analysis on the quality of the linear-programming (LP) relaxations. Table 4 reports the improvement of the objective values (%) and the percentages of fractional variables in the solutions of both the linear- and the root-relaxations. Both relaxations are obtained by removing the integrality restrictions from the original instances, however the root-relaxation also includes the cutting planes that CPLEX generated for the original MIPs, just before branching has been initiated. The two formulations do not differ significantly in terms of the objective function values of the relaxations. In particular, we observe negligible gain with the new formulation in the standard LP-relaxation, which tends to vanish in the root-relaxation (i.e., when the formulation is tightened with cutting planes). Although the differences of the objective values are small, remarkable differences appear in the percentages of fractional variables. In particular, the new formulation leads to solutions that contain significantly smaller numbers of fractional variables. The percentage of fractional variables never exceed 3%, and the root-relaxation leads to almost integral solutions. In contrast, under the base formulation,
the percentage of fractional variables can exceed 12% and 6% for the LP- and root-relaxations, respectively. These results favorably support the thesis that the new formulation has a stronger polyhedral structure, which is also consistent with the results on the number of B&B nodes reported in Table 3.

![Table 4: Improvement in the objective values of the LP- and root-relaxations with respect to the base formulation, and the percentage of fractional variables (|T| = 168).](image)

We highlight the improvements conveyed by each feature of the new formulation. Recall that the new formulation is achieved through introducing state-transition variables, and using improved versions of the piecewise-linear cost functions. Figure 3 displays the box plots of the computational times (in seconds) to solve the 100 week-long synthetic instances with 187 generators. It depicts (i) the base formulation, (ii) the new formulation without the improved versions of the piecewise-linear cost functions, and (iii) the new formulation with all of the suggested refinements. The box plots corroborate the conclusions of the computational analyses presented above. The new formulation performs better than the base formulation. All of the descriptive statistics, such as the measures of central tendency and variability, favor the new formulation. In particular, the new formulation without the improved piecewise-linear cost functions, allows us to reduce the average computational time by one third, from 196.3 to 63.7 seconds; which is further reduced by one half, from 63.7 to 30.8 seconds, by the use of the refined piecewise-linear cost functions.

Next, we present the computational results on the real-life instances. Specifically, we consider both a day-long and a week-long version of the instance, made accessible by Krall et al. (2012). The magnitude of these instances will provide better intuition on how the base and new formulations scale up to the industrial-strength problems. Indeed, in the real-life instances, the difference in the computational performance of the two formulations are more pronounced than our observations for the synthetic instances. For instance, we observe that the base formulation performs poorly when a time horizon of 168 hours (week-long) is considered, and cannot establish optimality within the allotted time limit of 7200 seconds. In contrast, the new formulation is able to establish optimality for the week-long instance in less than 10 minutes. The gain in the objective function values of the LP-relaxations due to the new formulation is considerably better than our previous observations on the synthetic instances. In particular, for the day-long instance we observe an improvement of 1.39%. Nevertheless, these improvements again diminish in the root-relaxations. Finally, we focus on the percentage of fractional variables. In contrast to the synthetic instances, the real-life instances lead to smaller numbers of fractional variables.
variables, and this further boosts the performance of the new formulation. In particular, we observe that the percentage of integral variables in the week-long instance at the root node reaches 99.76%, leading to an almost integer solution.

![Figure 3: Base vs New Formulation.](image)

<table>
<thead>
<tr>
<th>Branch-and-Bound</th>
<th>LP/Root Relaxation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cuts</td>
</tr>
<tr>
<td><strong>Base</strong></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>4,605</td>
</tr>
<tr>
<td>168</td>
<td>14,622</td>
</tr>
<tr>
<td><strong>New</strong></td>
<td>24</td>
</tr>
<tr>
<td>168</td>
<td>1,576</td>
</tr>
</tbody>
</table>

Table 5: Performance measures for the branch-and-bound algorithm and the LP relaxation of the studied real-life instance ($|G| = 939$)

*: Time until a loose relative optimality tolerance has been reached, namely 1%.

†: Aborted due to a time limit of 7200 seconds.

4. Conclusion

In this study, we developed a new formulation for the UC problem with the purpose of achieving better computational performance. The proposed formulation replaces the long-established state variables of the UC formulations with state-transition variables, and adopts production variables that only represent the variable portion of generation amounts. The reformulation of the constraints with the new variables leads to tighter descriptions of the UC polytopes. We compare the performance of the new formulation with a state-of-the-art formulation. Our results indicate that the new formulation performs significantly better in terms of solution times and numbers of branch-and-bound nodes. We further observe that using the new formulation, an industrial-strength instance with a week-long planning horizon can be solved in less than ten minutes. For the same instance, the benchmark formulation fails to establish optimality within two hours. This outcome provides clear support for the applicability of MIP-based algorithms to the real-life UC problems of the power industry,
and allows the consideration of more complex models with significantly less computational burden.

References


Appendix A  Facet Proofs

The minimum up/downtime inequalities and the ramping inequalities refer to each specific generator $g$. In what follows, we exploit this fact and drop the generator indices for notational brevity.

A.1 Proofs for the minimum up/downtime inequalities

Consider the following minimum uptime/downtime polytope:

$$
\mathcal{M} = \left\{ s, \tilde{x}, z \in \{0, 1\}^{\lvert T \rvert} : \begin{array}{l}
\sum_{i=t-UT+1}^{t-1} s_i \leq \tilde{x}_t \quad \forall t \in \{UT, \ldots, \lvert T \rvert\}, \\
\sum_{i=t-DT}^{t-1} s_i \leq 1 - \tilde{x}_{t-DT} \quad \forall t \in \{DT + 1, \ldots, \lvert T \rvert\} \\
\hspace{0.5cm} s_{t-1} + \tilde{x}_{t-1} = z_t + \tilde{x}_t \quad \forall t \in \{2, \ldots, \lvert T \rvert\}
\end{array} \right\}
$$

(12a)

(12b)

(9)

In the definition above, the history of the generators is neglected by removing the corresponding turn on and turn off constraints defined for the time periods $\{1, \ldots, UT - 1\}$, and $\{1, \ldots, DT\}$, respectively.

**Proposition 4.** The minimum up/downtime constraints (13) are facets of the polytope $\mathcal{M}$.

**Proof.** Observing that the equality constraints (9) are linearly independent, the dimension of the polytope $\mathcal{M}$ is at most $2 \times \lvert T \rvert + 1$. Indeed, $\dim(\mathcal{M}) = 2 \times \lvert T \rvert + 1$ because the following $2 \times \lvert T \rvert + 2$ integer solutions of $\mathcal{M}$ are affinely independent, which is easy to verify by ordering the solution in a matrix and placing first the $\tilde{x}$ variables and then the $s$ and the $z$ variables. We obtain a lower-triangular submatrix as shown below.

Consider a set of $|T|$ solutions $(s, \tilde{x}, z)$, $i \in \{1, \ldots, |T|\}$. Solution $(s, \tilde{x}, z)$, has the following structure. The components of the vector $s$ (resp. $z$) are set to zero with the only exception of the $i^{th}$ component (resp. $(i + UT + 1)^{th}$ or $|T|$, whichever is smaller) which is set to one. The components of $\tilde{x}$ are

$$
\tilde{x}_k = \begin{cases} 
1, & i + 1 \leq k \leq \min\{i + UT, |T|\} \\
0, & \text{otherwise}
\end{cases}
$$
|\{T\}| additional solutions \((s, \tilde{x}, z)\), \(j \in \{1 \ldots |T|\}\), which belong to a set \(J\), are obtained by setting the \(s\) and \(z\) vectors to null, while the vector \(\tilde{x}\) has the following components:

\[
\tilde{x}_k = \begin{cases} 
1, & k \leq j \\
0, & \text{otherwise} 
\end{cases}
\]

The last two solutions are \((s, \tilde{x}, z) = (0, \bar{0}, 0, \ldots, 0)\) and the null solution.

Given the dimension of the polytope \(\mathcal{M}\), any constraint (13a) is a facet if it has dimension \(2 \times |T|\), i.e., if there are \(2 \times |T| + 1\) affinely independent solutions which satisfy the inequality as equality. For some \(t \in \{UT, \ldots, |T|\}\), consider the set

\[
U_t = \{s, \tilde{x}, z \in \{0, 1\}^{|T|} : \sum_{i=t-UT+1}^{t-1} s_i = \tilde{x}_t \}.
\]

All the solutions listed above belong to the set \(U_t\), with the exception of \(|T| - (t + 1)\) solutions of set \(J\), which turn off the generator at time period \(t + 1\) or later, i.e., solutions \(j\)'s with \(j > t\). For each of these solutions, turn the generator on at time \(t - UT + 1\) which implies setting the \(t - UT + 1\) component of vector \(s\) to 1 and turning to 0 all the first \(t - UT + 1\) components of vector \(\tilde{x}\). \(2 \times |T| + 1\) affinely independent solutions have thus been generated, proving that \(U_t\) is a facet of \(\mathcal{M}\).

The proof for minimum downtime constraints (13b) is analogous and can be obtained by using the same arguments as in the proof for constraints (13a).

\section*{A.2 Proofs for the ramp up/down inequalities}

Here we ignore generator history and focus on the ramping inequalities defined only for \(t \in \{2, \ldots, |T|\}\).

\begin{proposition}
Assuming \(\bar{C} > \bar{R} + R \geq 2 \times R\), the ramp up/down inequalities, (14a) and (14b), and the generation limit constraints, (10a) and (12), are facets of the two-period ramping polytope.
\end{proposition}

\begin{proof}
The two-period ramping polytope \((R^2_t)\) is given below:

\[
R^2_t = \{(\bar{p}_t, r_{t-1}, r_t, s_t, \tilde{x}_t, z_t, t) \in \mathbb{R}^3_+ \times \{0, 1\}^3 : (10a), (12), (14a), (14b)\}
\]

The dimension of the two-period ramping polytope \((R^2_t)\) is at most 6. \tablename~6 displays feasible solutions of the \(R^2_t\) polytope. The first seven solutions are affinely independent thus proving the full dimensionality of the two-period ramping polytope.

Using the solutions displayed in Table 6, it is also easy to verify that constraints (10a), (12), (14a), (14b) are facet-defining for \(R^2_t\). Indeed, the last column of Table 6 reports the constraints which are not active in the corresponding solution. To prove that each of the constraints are facet-defining, modify the coefficients of the decision variables of the solutions in Table 6 as follows:

- constraint (10a), turn the coefficient of \(\bar{p}_t\) from \(\bar{S}\) to \(\bar{C}\) in solution 4.
- constraint (12), six solutions out of the seven already satisfy the constraint as equality.
- constraint (14a), turn the coefficients of \(r_{t-1}\) from \(\bar{C}\) to \(\bar{C} - \bar{R} - \bar{C}\) and from \(R\) to 0 in solutions 6 and 7 respectively.
- constraint (14b), replace solution 5 with solution 8 at the bottom of Table 6.

For each of the constraints, we have provided six affinely independent feasible solutions which satisfy all constraints as equality, thus proving the proposition.
\end{proof}

\begin{remark}
Observe that the state decision variables’ inequality \((s_t + \tilde{x}_t + z_t \leq 1)\) is also facet-defining for the two-period ramping polytope.
\end{remark}
An Improved MIP Formulation for the UC Problem

Atakan, S., Lulli, G., and Sen, S.

Table 6: Set of feasible solutions for the two-period ramping polytope.

Appendix B  Details on the Data Generation

B.1 Synthetic Instances

Extending the Time Horizon in the Synthetic Data  The instances studied in Ostrowski et al. (2012) consider a time horizon of 24 hours, in which the hourly demands are given as percentages of the total generation capacity of the system, and the hourly reserve amounts are set to 3% of the corresponding demand. To generate demands beyond the 24 hours, we utilized the demand pattern of the studied real-life instances (see Appendix B.2). In particular, for each hour $h \in \{1, \ldots, 24\}$ in day $d \geq 2$, we compute the ratio $\rho_{h,d} = \frac{d_{h,d}}{d_{h,1}}$, where $d_{h,1}$ is the demand in the winter test problem for time period $h$. Notice that $\rho_{h,d}$ simply calculates the ratio of the demand at hour $h$ of day $d$ to day 1. We multiplied these ratios with the demand percentages given in Ostrowski et al. (2012), and rounded the resulting values to the nearest integer. The resulting percentages are provided in Table 7.

<table>
<thead>
<tr>
<th>Hours</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71</td>
<td>68</td>
<td>66</td>
<td>64</td>
<td>63</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td>2</td>
<td>65</td>
<td>63</td>
<td>60</td>
<td>58</td>
<td>58</td>
<td>57</td>
<td>58</td>
</tr>
<tr>
<td>3</td>
<td>62</td>
<td>60</td>
<td>58</td>
<td>55</td>
<td>55</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>59</td>
<td>56</td>
<td>54</td>
<td>54</td>
<td>53</td>
<td>52</td>
</tr>
<tr>
<td>5</td>
<td>58</td>
<td>58</td>
<td>55</td>
<td>52</td>
<td>53</td>
<td>52</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>58</td>
<td>60</td>
<td>57</td>
<td>54</td>
<td>56</td>
<td>54</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>66</td>
<td>63</td>
<td>59</td>
<td>62</td>
<td>59</td>
<td>52</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
<td>71</td>
<td>68</td>
<td>65</td>
<td>67</td>
<td>64</td>
<td>56</td>
</tr>
<tr>
<td>9</td>
<td>73</td>
<td>80</td>
<td>76</td>
<td>74</td>
<td>75</td>
<td>73</td>
<td>65</td>
</tr>
<tr>
<td>10</td>
<td>80</td>
<td>86</td>
<td>83</td>
<td>81</td>
<td>81</td>
<td>81</td>
<td>75</td>
</tr>
<tr>
<td>11</td>
<td>82</td>
<td>87</td>
<td>86</td>
<td>84</td>
<td>83</td>
<td>84</td>
<td>79</td>
</tr>
<tr>
<td>12</td>
<td>83</td>
<td>88</td>
<td>87</td>
<td>85</td>
<td>84</td>
<td>87</td>
<td>83</td>
</tr>
</tbody>
</table>

Table 7: Hourly demands, given as percentages of the total generation capacity of the system (%).

Additional Instances  Among the week-long instances described above, we have picked the one with the largest number of generators ($|\mathcal{G}| = 187$) and randomly generated additional instances. The data generation scheme resembles Ostrowski et al. (2012), where the individual generators are randomly drawn out of the 8 generator types given in the paper. More specifically, until 187 generators are selected, we randomly pick a generator type according to a multinomial distribution, where the probability of each generator type is given in Table 8. These probabilities correspond to the capacity ratio of a type $k$ generator to the sum of capacities of each generator type. The rest of the data generation follows Ostrowski et al. (2012).
### Generator Parameters

<table>
<thead>
<tr>
<th>Generator Type</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.293</td>
<td>0.293</td>
<td>0.084</td>
<td>0.084</td>
<td>0.104</td>
<td>0.052</td>
<td>0.055</td>
<td>0.035</td>
</tr>
</tbody>
</table>

Table 8: The parameters of the multinomial distribution for random generator selection.

#### B.2 Real-Life Instances

The real-life instances of §3 are based on the winter test problem of Krall et al. (2012), available through the FERC website\(^1\). The details of the generation of the original data is documented in Krall et al. (2012), therefore we only give our modifications and some clarifications. We considered the “seasonal capability” of the generators as their capacities. The start up and shut down rates are assumed to be 70% of the ramp up and ramp down rates, respectively. The minimum generation requirements are either set to their given values, or to the start up and shut down rates, whichever is the minimum. If these amounts are not available or given as 0, we set them to 1. Similarly, missing minimum uptime/downtime values are set to 1. Generators with no cost curves or containing invalid entries are removed from consideration (a total of 104 generators were removed). Only the initial on/off statuses of the generators are provided in the original data. For ease of replicability, we assumed that the minimum uptime/downtime requirements are not restrictive at the initial time period. Similar to the data in Ostrowski et al. (2012), the fixed costs are given as hot and cold start up costs. We assumed that all generators cool off in 5 time units. Finally, the demand data in Krall et al. (2012) consists of only 24 hours. To be consistent in generating longer time horizons, we ignored the provided demand data and extracted the realized demand from the PJM website\(^2\). More specifically, we considered the demand in the PJM region between 01/31/2010 and 01/06/2010, where the former marks the date that the original data is based on.

\(^{1}\)http://www.ferc.gov/industries/electric/indus-act/market-planning/rto-commit-test.asp